Optimization

Lectured by Dr Dante Kalise

Typed by Aris Zhu Yi Qing

April 25, 2022

Contents

Chapter 1

Mathematical Preliminaries

1.1 Topological Concepts

Definition 1. The **open ball** with center $c \in \mathbb{R}^n$ and radius r is

$$B(c, r) = \left\{ \mathbf{x} : \|\mathbf{x} - c\| < r \right\}.$$

Similarly, the **closed ball** with center c and radius r is

$$B[c, r] = \left\{ \mathbf{x} : \|\mathbf{x} - c\| \leq r \right\}.$$

Definition 2. Given a set $U \subseteq \mathbb{R}^n$, a point $\mathbf{c} \in U$ is called an **interior point** of U if $\exists r > 0$ for which $B(\mathbf{c}, r) \subseteq U$. The set of all interior points of a given set U is called the interior of the set and is denoted by

$$\operatorname{int}(U) = \left\{ \mathbf{x} \in U : B(\mathbf{x}, r) \subseteq U \text{ for some } r > 0 \right\}.$$

Definition 3. Given a set $U \subseteq \mathbb{R}^n$, a **boundary point** of U is a vector $\mathbf{x} \in \mathbb{R}^n$ satisfying that any neighbourhood of $\mathbf{x}$ contains at least one point in U and at least one point in its completement U^c. We denote

$$\operatorname{bd}(U) = \text{The set of all boundary points of a set } U.$$

Definition 4. The **closure** of a set $U \subseteq \mathbb{R}^n$ is the smallest closed set containing U, denoted by $\operatorname{cl}(U)$ with

$$\operatorname{cl}(U) = U \cup \operatorname{bd}(U).$$

Definition 5. A set $U \subseteq \mathbb{R}^n$ is called **bounded** if $\exists M > 0$ for which $U \subseteq B(0, M)$.

Definition 6. A set $U \subseteq \mathbb{R}^n$ is called **compact** if it is closed and bounded.

1.2 Multi-variable Calculus

Definition 7. The **directional derivative** of a scalar function f w.r.t. $\mathbf{d}$ at a point $\mathbf{x}$ is denoted as

$$f'(\mathbf{x};\mathbf{d}) = \nabla f(\mathbf{x})^T\mathbf{d}$$

Theorem 8. Given the general quadratic functions of the form

$$f(\mathbf{w}) = \mathbf{w}^T A\mathbf{w} + \mathbf{b}^T\mathbf{w} + \gamma$$

we have

$$\nabla f(\mathbf{w}) = (A^T + A)\mathbf{w} + \mathbf{b}, \qquad \nabla^2 f(\mathbf{w}) = A + A^T.$$

If A is symmetric, then

$$\nabla f(\mathbf{w}) = 2A\mathbf{w} + \mathbf{b}, \qquad \nabla^2 f(\mathbf{w}) = 2A.$$

1.3 Positive Definiteness of Matrix

Proposition 9. Let A be a positive definite (semidefinite) matrix, then

- the diagonal elements of A are positive (nonnegative)
- $\mathrm{Tr}(A)$ and $\det(A)$ are positive (nonnegative)

(Test 1) Theorem 10. Let $A \in \mathbb{R}^{n\times n}$ be symmetric, then

- A is positive definite (semidefinite) iff all its eigenvalues are positive (nonnegative).
- A is indefinte iff it has at least one positive eigenvalue and at least one negative eigenvalue.

Definition 11. Let $A \in \mathbb{R}^{n\times n}$ be symmetric, then

- A is **diagonally dominant** if

$$|A_{ii}| \geq \sum_{j\neq i} |A_{ij}| \quad \forall i = 1, 2, \ldots, n$$

- A is **strictly diagonally dominant** if

$$|A_{ii}| > \sum_{j\neq i} |A_{ij}| \quad \forall i = 1, 2, \ldots, n$$

(Test 2) Theorem 12. If $A \in \mathbb{R}^{n \times n}$ is symmetric, diagonally dominant with positive (nonnegative) diagonal elements, then A is positive definite (semidefinite).

Chapter 2

Unconstrained Optimization

2.1 Optimums

Definition 13. Let $f : S \to \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$, then $\forall \mathbf{x} \in S$,

$\mathbf{x}^* \in S$ is a **global minimum** point of f over S if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$,

$\mathbf{x}^* \in S$ is a **strict global minimum** point of f over S if $f(\mathbf{x}) > f(\mathbf{x}^*)$,

and similar definitions for maximum.

Definition 14. Let $f : S \to \mathbb{R}$ be defined on a set $S \subseteq \mathbb{R}^n$, $\mathbf{x}^* \in S$ is a **local minimum** of f over S if $\exists r > 0$ s.t. $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for any $\mathbf{x} \in S \cap B(\mathbf{x}^*, r)$. Similar definitions for **strict local minimum** and maximum.

Definition 15. Let $f : U \to \mathbb{R}$ be a function defined on a set $U \subseteq \mathbb{R}^n$. Suppose that $\mathbf{x}^* \in \text{int}(U)$ and that all the partial derivatives of f are defined at $\mathbf{x}^*$, then $\mathbf{x}^*$ is called a **stationary point** of f if $\nabla f(\mathbf{x}^*) = 0$.

2.2 Second-order Optimality Conditions

Theorem 16. Let $f : U \to \mathbb{R}$ be a function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that $\mathbf{x}^*$ is a stationary point, then

- $\mathbf{x}^*$ is a local minimum point $\iff \nabla^2 f(\mathbf{x}^*) \succeq 0$.
- $\mathbf{x}^*$ is a strict local minimum point $\iff \nabla^2 f(\mathbf{x}^*) \succ 0$.
- similar necessary and sufficient conditions for (strict) local maximum point

Definition 17. Let $f : U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. A stationary point $\mathbf{x}^* \in U$ is called a **saddle point** of f over U if it is neither a local minimum nor a local maximum point of f over U.

Theorem 18. Let $f : U \to \mathbb{R}$ be a continuously differentiable function defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that f is twice continuously differentiable over U and that $\mathbf{x}^*$ is a stationary point. Then

$$\nabla^2 f(\mathbf{x}^*) \text{ is an indefinite matrix} \Longrightarrow \mathbf{x}^* \text{ is a saddle point of } f \text{ over } U.$$

2.3 Attainment of Minimal/Maximal Points

(Weierstrass') Theorem 19. Let f be a continuous function defined over a nonempty conpact set $C \subseteq \mathbb{R}^n$. Then $\exists$ a global minimum point of f over C and a global maximum point of f over C.

Definition 20. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function over $\mathbb{R}^n$. f is called **coercive** if

$$\lim_{\|\mathbf{x}\| \to \infty} f(\mathbf{x}) = \infty$$

Theorem 21. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous and coercive function and let $S \subseteq \mathbb{R}^n$ be a nonempty closed set. Then f attains a global minimum point on S.

2.4 Global Optimality Conditions

Theorem 22. Let f be a twice continuously differentiable function defined over $\mathbb{R}^n$. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a stationary point of f. Then

$$\nabla^2 f(\mathbf{x}) \succeq 0 \ \forall \mathbf{x} \in \mathbb{R}^n \Longrightarrow \mathbf{x}^* \text{ is a global minimum point of } f.$$

Proposition 23. Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$, with $A \in \mathbb{R}^{n\times n}$ symmetric, then

1. $\mathbf{x}$ is a stationary point of f iff $A\mathbf{x} = -\mathbf{b}$.
2. if $A \succeq 0$, then $\mathbf{x}$ is a global minimum point of f iff $A\mathbf{x} = -\mathbf{b}$.
3. if $A \succ 0$, then $\mathbf{x} = -A^{-1}\mathbf{b}$ is a strict global minimum point of f.

Chapter 3

Linear Least Squares

3.1 Problem Formulation

Consider the linear system

$$S\mathbf{x} \approx \mathbf{b}, \quad (S \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, m > n)$$

To solve the above system, the usual approach is to transform it to become

$$\min_{\mathbf{x}} \|S\mathbf{x} - \mathbf{b}\|^2 \iff \min_{\mathbf{x}\in\mathbb{R}^n} \left\{ f(\mathbf{x}) \equiv \mathbf{x}^T S^T S\mathbf{x} - 2\mathbf{b}^T S\mathbf{x} + \|\mathbf{b}\|^2 \right\}.$$

Note that $\nabla^2 f(\mathbf{x}) = 2S^T S \succeq 0$ since $\mathbf{x}^T S^T S\mathbf{x} = (S\mathbf{x})^T(S\mathbf{x}) = \|S\mathbf{x}\|^2 \geq 0$. Therefore, the unique optimal solution $\mathbf{x}_{\text{LS}}$ is the solution $\nabla f(\mathbf{x}) = 0$, namely

$$(S^T S)\mathbf{x}_{\text{LS}} = S^T\mathbf{b} \Longrightarrow \mathbf{x}_{\text{LS}} = \left(S^T S\right)^{-1} S^T\mathbf{b}.$$

3.2 Data Fitting

1. For dataset $(\mathbf{s}_i, b_i)$ where $\mathbf{s}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$, we could transform to problem

$$\min_{\mathbf{x}} \sum_{i=1}^{m} \left(\mathbf{s}_i^T\mathbf{x} - b_i\right)^2 \Longrightarrow \min_{\mathbf{x}} \|S\mathbf{x} - \mathbf{b}\|^2$$

2. For polynomial fitting, given a set of points $\mathbb{R}^2 : (u_i, y_i)$, the associated linear system is

$$\begin{pmatrix} 1 & u_1 & u_1^2 & \cdots & u_1^d \\ 1 & u_2 & u_2^2 & \cdots & u_2^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_m & u_m^2 & \cdots & u_m^d \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{pmatrix}$$

3.3 Regularized Least Squares

A Regularized Least Square problem is formulated as

$$\min_{\mathbf{x}} \|S\mathbf{x} - \mathbf{b}\|^2 + \lambda R(\mathbf{x}),$$

where λ is the regularization parameter and $R(\cdot)$ is the regularization function (also called a *penalty* function). A common choice is a quadratic regularization function:

$$\min_{\mathbf{x}} \|S\mathbf{x} - \mathbf{b}\|^2 + \lambda \|D\mathbf{x}\|^2$$

with its optimal solution being

$$\mathbf{x}_{\text{RLS}} = (S^T S + \lambda D^T D)^{-1} S^T \mathbf{b}$$

since $\nabla f = 2S^T S\mathbf{x} - 2S^T\mathbf{b} + 2\lambda D^T D\mathbf{x} = 0$.

3.4 Denoising

Suppose a noisy measurement of a signal $\mathbf{x} \in \mathbb{R}^n$ is given

$$\mathbf{b} = \mathbf{x} + \mathbf{w}$$

where $\mathbf{x}$ is the "true" unknown signal, $\mathbf{w}$ is the unknown noise and $\mathbf{b}$ is the (known) measures vector. We could define

$$R(\mathbf{x}) = \|L\mathbf{x}\|^2, \text{ where } L = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}$$

as the regularization function to penalize any sudden variations in signal. The RLS is thus

$$\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{b}\|^2 + \lambda \|L\mathbf{x}\|^2$$

with its direct solution being

$$\mathbf{x}_{\text{RLS}}(\lambda) = (I + \lambda L^T L)^{-1} \mathbf{b}.$$

Chapter 4

The Gradient Method